Lecture 6

Last updated by Serik Sagitov: December 3, 2012

Abstract

A course based on the book Probability and Random Processes by Geoffrey Grimmett and David Stirzaker. Chapter 7.4-7.9. Convergence of random variables. Chapter 8. Random processes.

1 Inequalities

Jensen's inequality. Given a convex function $J(x)$ and a random variable X with mean μ we have

$$\mathbb{E}(J(X)) \geq J(\mu).$$

Proof. Due to convexity there is λ such that $J(x) \geq J(\mu) + \lambda(x - \mu)$. Thus

$$\mathbb{E}(J(X)) \geq \mathbb{E}(J(\mu) + \lambda(X - \mu)) = J(\mu).$$

Markov's inequality. For any random variable X and $a > 0$

$$\mathbb{P}(|X| > a) \leq \frac{\mathbb{E}|X|}{a}.$$

Proof:

$$\mathbb{E}|X| \geq \mathbb{E}(|X| I_{\{|X|>a\}}) \geq a\mathbb{E}(I_{\{|X|>a\}}) = a\mathbb{P}(|X| > a).$$

Chebyshev's inequality. Given a random variable X with mean μ and variance σ^2 for any $\epsilon > 0$ we have

$$\mathbb{P}(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

Proof:

$$\mathbb{P}(|X - \mu| > \epsilon) = \mathbb{P}((X - \mu)^2 > \epsilon^2) \leq \frac{\mathbb{E}((X - \mu)^2)}{\epsilon^2}.$$

Cauchy-Schwartz's inequality: for r.v. X and Y

$$\big(\mathbb{E}(XY)\big)^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$$

with equality if only if $aX + bY = 1$ a.s. for some non-trivial pair of constants (a, b).
Hölder's inequality. If $p, q > 1$ and $p^{-1} + q^{-1} = 1$, then

$$\mathbb{E}|XY| \leq \big(\mathbb{E}|X^p|\big)^{1/p}\big(\mathbb{E}|Y^q|\big)^{1/q}.$$

Minkowski's inequality. Triangle inequality. If $p \geq 1$, then

$$\big(\mathbb{E}|X + Y|^p\big)^{1/p} \leq \big(\mathbb{E}|X^p|\big)^{1/p} + \big(\mathbb{E}|Y^p|\big)^{1/p}.$$

Kolmogorov's inequality. Let $\{X_n\}$ be iid with zero means and variances σ_n^2. Then for any $\epsilon > 0$

$$\mathbb{P}(\max_{1\leq i\leq n} |X_1 + \ldots + X_i| > \epsilon) \leq \frac{\sigma_1^2 + \ldots + \sigma_n^2}{\epsilon^2}.$$

Doob-Kolmogorov's inequality. If $\{S_n\}$ is a martingale, then for any $\epsilon > 0$

$$\mathbb{P}(\max_{1\leq i\leq n} |S_i| > \epsilon) \leq \frac{\mathbb{E}(S_n^2)}{\epsilon^2}.$$

2 Strong LLN

Theorem 1 *Let $X_1, X_2, \ldots$ be iid random variables defined on the same probability space with mean μ and finite second moment. Then*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{L^2} \mu.$$

Proof. Since $\sigma^2 := \mathbb{E}(X_1^2) - \mu^2 < \infty$, we have

$$\mathbb{E}\left(\left(\frac{X_1 + \ldots + X_n}{n} - \mu\right)^2\right) = \mathbb{V}ar\left(\frac{X_1 + \ldots + X_n}{n}\right) = \frac{n\sigma^2}{n^2} \to 0.$$

Theorem 2 *Strong LLN. Let $X_1, X_2, \ldots$ be iid random variables defined on the same probability space. Then*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{a.s.}} \mu$$

for some constant μ iff $\mathbb{E}|X_1| < \infty$. In this case $\mu = \mathbb{E}X_1$ and $\frac{X_1+\ldots+X_n}{n} \xrightarrow{L^1} \mu$.

There are cases when convergence in probability holds but not a.s. In those cases of course $\mathbb{E}|X_1| = \infty$.

Theorem 3 *The law of the iterated logarithm. Let $X_1, X_2, \ldots$ be iid random variables with mean 0 and variance 1. Then*

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{X_1 + \ldots + X_n}{\sqrt{2n \log\log n}} = 1\right) = 1$$

and

$$\mathbb{P}\left(\liminf_{n\to\infty} \frac{X_1 + \ldots + X_n}{\sqrt{2n \log\log n}} = -1\right) = 1.$$

Proof*. The second assertion follows from the first one after applying it to $-X_i$. The proof of the first part is difficult. One has to show that the events

$$A_n(c) = \{X_1 + \ldots + X_n \geq c\sqrt{2n \log\log n}\}$$

occur for infinitely many values of n if $c < 1$ and for only finitely many values of n if $c > 1$.

3 Martingales

Martingale: a betting strategy. Let X_n be the gain of a gambler doubling the bet after each loss. The game stops after the first win.

- $X_0 = 0$
- $X_1 = 1$ with probability $1/2$ and $X_1 = -1$ with probability $1/2$,
- $X_2 = 1$ with probability $3/4$ and $X_2 = -3$ with probability $1/4$,
- $X_3 = 1$ with probability $7/8$ and $X_3 = -7$ with probability $1/8, \ldots,$
- $X_n = 1$ with probability $1 - 2^{-n}$ and $X_n = -2^n + 1$ with probability 2^{-n}.

Conditional expectation

$$\mathbb{E}(X_{n+1}|X_n) = (2X_n - 1)\frac{1}{2} + (1)\frac{1}{2} = X_n.$$

If N is the number of games, then $\mathbb{P}(N = n) = 2^{-n}$, $n = 1, 2, \ldots$ with $\mathbb{E}(N) = 2$ and

$$\mathbb{E}(X_{N-1}) = \mathbb{E}(1 - 2^{N-1}) = 1 - \sum_{n=1}^{\infty} 2^{n-1}2^{-n} = -\infty.$$

Definition 4 *The sequence $\{S_n\}_{n\geq 1}$ is a martingale with respect to the sequence $\{X_n\}_{n\geq 1}$, if for all $n \geq 1$*

- $\mathbb{E}|S_n| < \infty$,
- $\mathbb{E}(S_{n+1}|X_1, \ldots, X_n) = S_n$.

It follows that $S_n = \psi_n(X_1, \ldots, X_n)$. We sometimes just say that $\{S_n\}_{n\geq 1}$ is a martingale.

A submartingale. If $\{S_n\}_{n\geq 1}$ is a martingale, then $S_{n+1} - S_n$ and S_n are uncorrelated since

$$\mathbb{E}(S_n(S_{n+1} - S_n)|X_1, \ldots, X_n) = 0.$$

It follows that $\{S_n^2\}_{n\geq 1}$ is a submartingale

$$\begin{aligned}\mathbb{E}(S_{n+1}^2|X_1, \ldots, X_n) &= \mathbb{E}((S_{n+1} - S_n)^2 + 2S_n(S_{n+1} - S_n) + S_n^2|X_1, \ldots, X_n) \\ &= \mathbb{E}((S_{n+1} - S_n)^2|X_1, \ldots, X_n) + S_n^2 \geq S_n^2.\end{aligned}$$

We have

$$\mathbb{E}(S_{n+1}^2) = \mathbb{E}(S_n^2) + \mathbb{E}((S_{n+1} - S_n)^2)$$

so that $\mathbb{E}(S_n^2)$ is non-decreasing and there always exists a finite or infinite limit

$$M = \lim_{n\to\infty} \mathbb{E}(S_n^2). \tag{1}$$

More generally due to the Jensen inequality: if $\{S_n\}_{n\geq 1}$ is a martingale and $J(x)$ is convex, then $\{J(S_n)\}_{n\geq 1}$ is a submartingale.

Example. A simple random walk $S_n = X_1 + \ldots + X_n$ with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = -1) = q$. In the symmetric case $p = q$

- S_n is a martingale: $\mathbb{E}(S_{n+1}|X_1, \ldots, X_n) = S_n + \mathbb{E}(X_{n+1}) = S_n$,
- $S_n^2 - n$ is a martingale: $\mathbb{E}(S_{n+1}^2 - n - 1|X_1, \ldots, X_n) = S_n^2 + 2S_n\mathbb{E}(X_{n+1}) + \mathbb{E}(X_{n+1}^2) - n - 1 = S_n^2 - n$.

De Moivre's martingale $D_n = (q/p)^{S_n}$:

$$\mathbb{E}(D_{n+1}|X_1, \ldots, X_n) = p(q/p)^{S_n+1} + q(q/p)^{S_n-1} = (q/p)^{S_n} = D_n.$$

Theorem 5 *If $\{S_n\}$ is a martingale with finite M defined by (1), then there exists a random variable S such that $S_n \to S$ a.s. and in mean square.*

Proof*. Step 1. Put

$$A_m(\epsilon) = \bigcup_{i\geq 1}\{|S_{m+i} - S_m| \geq \epsilon\}.$$

Using the Doob-Kolmogorov inequality show that $\mathbb{P}(A_m(\epsilon)) \to 0$ as $m \to \infty$ for any $\epsilon > 0$.
Step 2. Show that the sequence $\{S_n\}$ is a.s. Cauchy convergent:

$$\mathbb{P}\Big(\bigcap_{\epsilon>0}\bigcup_{m\geq 1} A_m^c(\epsilon)\Big) \to 0$$

which implies the existence of S such that $S_n \to S$ a.s.
Step 3. Prove the convergence in mean square using the Fatou lemma

$$\begin{aligned}\mathbb{E}((S_n - S)^2) &= \mathbb{E}(\liminf_{m\to\infty}(S_n - S_m)^2) \leq \liminf_{m\to\infty} \mathbb{E}((S_n - S_m)^2) \\ &= \liminf_{m\to\infty} \mathbb{E}(S_m^2) - \mathbb{E}(S_n^2) = M - \mathbb{E}(S_n^2) \to 0, \quad n \to \infty.\end{aligned}$$

4 Predictions and conditional expectation

Definition 6 *Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}(Y^2) < \infty$. The best predictor of Y given the knowledge of X is the function $\hat{Y} = h(X)$ that minimizes $\mathbb{E}((Y - \hat{Y})^2)$.*

Let $L^2(\Omega, \mathcal{F}, \mathbb{P})$ be the set of random variables Z on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}(Z^2) < \infty$. Define a scalar product on the linear space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ by $\langle U, V \rangle = \mathbb{E}(UV)$ leading to the norm

$$\|Z\| = \langle Z, Z \rangle^{1/2} = (\mathbb{E}(Z^2))^{1/2}.$$

Let H be the subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ of all functions of X having finite second moment

$$H = \{h(X) : \mathbb{E}(h(X)^2) < \infty\}.$$

Geometrically, the best predictor of Y given X is the projection $\hat{Y}$ of Y on H so that

$$\mathbb{E}((Y - \hat{Y})Z) = 0, \text{ for all } Z \in H. \tag{2}$$

Theorem 7 *Let X and Y be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{E}(Y^2) < \infty$. The best predictor of Y given X is the conditional expectation $\hat{Y} = \mathbb{E}(Y|X)$.*

Proof. Put $\hat{Y} = \mathbb{E}(Y|X)$. We have due to the Jensen inequality $\hat{Y}^2 \le \mathbb{E}(Y^2|X)$ and therefore

$$\mathbb{E}(\hat{Y}^2) \le \mathbb{E}(\mathbb{E}(Y^2|X)) = \mathbb{E}(Y^2) < \infty.$$

It remains to verify (2):

$$\mathbb{E}((Y - \hat{Y})Z) = \mathbb{E}(\mathbb{E}(Y - \hat{Y})Z|Z)) = \mathbb{E}(\mathbb{E}(Y|X)Z - \hat{Y}Z) = 0.$$

To prove uniqueness assume that there is another predictor $\bar{Y}$ with $\mathbb{E}((Y - \bar{Y})^2) = \mathbb{E}((Y - \hat{Y})^2) = d^2$. Then $\mathbb{E}((Y - \frac{\hat{Y}+\bar{Y}}{2})^2) \ge d^2$ and according to the parallelogram rule

$$2\Big(\|Y - \hat{Y}\|^2 + \|Y - \bar{Y}\|^2\Big) = 4\|Y - \frac{\hat{Y} + \bar{Y}}{2}\|^2 + \|\bar{Y} - \hat{Y}\|^2$$

we have

$$\|\bar{Y} - \hat{Y}\|^2 \le 2\Big(\|Y - \hat{Y}\|^2 + \|Y - \bar{Y}\|^2\Big) - 4d^2 = 0.$$